Complexity of Synthetic Reactions. The Use of Complexity Indices to Evaluate Reactions, Transforms and Disconnections. (Supplementary Information)

Steven H. Bertz

Complexity Study Center, Mendham, NJ 07945 USA

This submission was created using the RSC Article Template

Prolegomenon

A *graph* G consists of a finite set $V(G)$ of *points* (or *vertices*) together with a finite set $E(G)$ of *lines* (or *edges*), which are unordered pairs of distinct points of $V(G)$. When the lines of a graph are *arcs* (or *directed lines*), which are ordered pairs of distinct points, it is a *directed graph* or *digraph*. The *null graph* G_0 has the null set ϕ for $V(G)$ and $E(G)$. A line $x = uv = vu$ in G *joins* points u and v, which are *adjacent points*. Point u and line x are *incident* to each other. The *degree* d of a point is the number of lines incident to it. Two lines that share a point are *adjacent lines*, e.g., uv and vw. In a *multigraph* more than one line, i.e., a *multiple line*, joins at least one pair of points. A *loop* $x = uu$ is a line that joins a point to itself. A *pseudograph* allows both multiple lines and loops. The *union* $G_1 \cup G_2$ of graphs G_1 and G_2 has $V = V_1 \cup V_2$ and $E = E_1 \cup E_2$. *Disjoint graphs* G_1 and G_2 have no points (and hence no lines) in common, i.e., $V_1 \cap V_2 = \phi$ and $E_1 \cap E_2 = \phi$.

A *path* P_n is a sequence of points $p_1, p_2, p_3, \ldots, p_n$ that are joined by lines $l_1l_2, l_2l_3, \ldots, l_{n-1}l_n$. In a *proper* (or *self-avoiding*) *path* each point is distinct. (Hence each line is also distinct.) *Line-disjoint paths* can share points, but not lines. In a *connected graph* (or *1-component graph*) all pairs of points are the endpoints of some path. The length of the longest proper path is the *diameter* D of the graph. A *cycle* (or *ring*) C_n is a sequence of points $p_1, p_2, p_3, \ldots, p_n$ such that the first one p_1 and last one p_n are joined by line p_1p_n, and all n points are distinct. A *tree* is a connected graph without cycles (including loops and multiple lines).

The n points in a *labeled graph* are distinguished from each other by unique labels $\lambda_1, \lambda_2, \ldots, \lambda_n$. Alternatively, more than one point can have the same label. To avoid confusion these non-unique labels are called *colors*, and to assign them is *to color* the points. A *coloring* of a graph assigns colors to its points in such a way that no two adjacent points have the same color. A *point-labeled graph* has at least one labeled (e.g., colored) point that is different from the rest, and analogously for *line-labeled graph*. A *bipartite graph* or *bigraph* is a graph that has a 2-coloring, i.e., every line joins a point of the first color to a point of the second. All trees are bigraphs, including *star graphs* (or simply *stars*) $K_{1,n-1}$ on n points, where one point of the first color is joined to $n - 1$ points of the second.

Two graphs G and H are *isomorphic*, $G \cong H$, if and only if there exists a one-to-one correspondence between their point sets that preserves adjacency. Thus, it does not matter when a graph is relabeled, redrawn or manipulated in any way that leaves the adjacency relation intact; the result is the same graph. An *invariant* of graph G is a number $I(G)$ associated with G that has the same value for any graph H isomorphic to G; thus, $I(G) = I(H)$ whenever $G \cong H$. Examples of invariants are the number of points n, the number of lines e and the number of pairs of adjacent lines (connections) η. A *topological index* is a real number that is a graph invariant or is derived from one or more invariants.

Two graphs X and Y are *homeomorphic* if and only if they can be derived from a third graph Z by the subdivision of lines. A line $x = uv$ is subdivided by replacing it with two lines uw and wv that are adjacent at a new point w, which has $d = 2$. An equivalent definition is based on excision of a point w of degree 2 and replacement of lines uw and wv with a single line uv. A graph is *planar* if and only if it has no subgraph homeomorphic to K_5 or $K_{3,3}$.

A *subgraph* $S_i(G)$ of graph G is a graph that has all its points in $V(G)$ and lines in $E(G)$. Based on this definition, we include G itself in the set of all possible subgraphs. A *spanning subgraph* of G is a subgraph containing all the points of G. A spanning subgraph that is also a tree is a *spanning tree*. For any subset S of $V(G)$, the *subgraph of G induced by S* (or *induced subgraph* $\langle S \rangle$) is the maximal subgraph of G with point set S, i.e., two points are adjacent in $\langle S \rangle$ if and only if they are adjacent in G. If G_1 is a subgraph of G_2, then G_2 is a *supergraph* of G_1. A c-component graph comprises c disjoint, connected subgraphs.

In a *complete graph* K_n on n points, each point is joined to every other one, i.e., all pairs of points are adjacent. A *clique* in G is a complete subgraph of G. Examples of common cliques in molecular graphs are K_1 (methane), K_2 (ethane), $K_3 \cong C_3$ (cyclopropane) and K_4 (tetrahedrane). A *complete bipartite graph* or *complete bigraph* $K_{m,n}$ contains m points of one color, n points of a second and all possible mn lines. A *biclique* in G is a complete bipartite subgraph of G. Examples of common bicliques in molecular graphs include $K_{2,2} \cong C_4$ (cyclobutane) and the star graphs $K_{1,2}$, $K_{1,3}$ and $K_{1,4}$, which represent methylene, methine and quaternary carbon atoms, respectively.

A *homologous series* $\{H\}$ is a recursively generated sequence of connected graphs $G_1, G_2, G_3, \ldots$, where G_{i+1} is obtained by incrementing G_i according to a *recurrent rule*, which is a set of instructions that is repeated at each stage. Chemically relevant examples of homologous series are paths $\{P_n\}$, cycles $\{C_n\}$, star graphs $\{K_{1,n-1}\}$ and complete graphs $\{K_n\}$.

Molecules are *isomers* whenever they have the same number and kinds of atoms (i.e., the same molecular formula) and the same number of bonds. By analogy, two point-colored graphs are *isomers* (or *isomeric graphs*) whenever they have the same number of points of each color and the same number of edges.

The *line graph* $L(G)$ of a graph G has the lines of G as its points, and two points in $L(G)$ are adjacent whenever the corresponding lines in G are adjacent. Thus, a pair of adjacent lines in G, called a *connection*, becomes a line in $L(G)$. When counting adjacent lines in pseudographs, a loop is not adjacent to itself and a double line has one adjacency. Repeating the process leads to the *iterated line graphs* $L^i(G)$ ($i = 0, 1, 2, \ldots$), where $L^0(G) \cong G$ and $L^1(G) \cong L(G)$. For completeness, $L(C_1) \cong P_1$, $L(P_1) \cong P_0 \cong G_0$ and $L(G_0)$ is not defined. The *iterated line graph sequence* is the series of integers N^i, which are the numbers of lines in the iterated line graphs.

An *edge cover* $\mathfrak{C}$ of graph G is any family $\mathfrak{C} = \{S_1, \ldots, S_n\}$ of subgraphs $S_i(G)$ such that every edge of G is contained in $E(S_i)$ for some i. In an *edge clique cover* every subgraph $S_i \in \mathfrak{C}$ is a clique in G; in an *edge biclique cover* every $S_i \in \mathfrak{C}$ is a biclique in G. A *minimal edge cover* does not properly contain any other edge cover. Thus, in a minimal edge cover each $S_i \in \mathfrak{C}$ is essential, i.e., S_i contains at least one edge of G that is not in any other subgraph $S_j \in \mathfrak{C}$ ($i \neq j$). A *partition* P is an edge cover with the additional property that each edge belongs to exactly one $S_i \in \mathfrak{C}$. Thus, in a partition no two subgraphs S_i and S_j ($i \neq j$) have an edge in common.

Table 13 Selected index values for all possible disconnections of azacyclohexane (**28**).[a]

disconnection (**28** ⇒)	N_T(lpe)	$-\Delta N_T$(lpe)	N_S(lpe)	$-\Delta N_S$(lpe)	$C(\eta,\varepsilon)$	$-\Delta C(\eta,\varepsilon)$
28	59	0	31	0	30.92	0.0
one-bond						
29	27	32	17	14	19.90	11.0
30	31	28	22	9	19.90	11.0
31	33	26	21	10	19.90	11.0
two-bond						
32	17	42	7	24	7.51	23.4
33	21	38	14	17	13.12	17.8
34	24	35	17	14	13.12	17.8
35	25	34	14	17	11.12	19.8
36	15	44	8	23	4.00	26.9
37	17	42	11	20	7.24	23.7
38	19	40	12	19	7.24	23.7
39	15	44	9	22	2.75	28.2
40	16	43	9	22	2.75	28.2
three-bond						
41	13	46	6	25	2.00	28.9
42	16	43	11	20	7.24	23.7
43	18	41	12	19	7.24	23.7
44	11	48	5	26	0.00	30.9
45, **46**[b]	12	47	7	24	2.00	28.9
47, **48**[b]	13	46	8	23	2.75	28.2
49	14	45	8	23	2.75	28.2
50	11	48	6	25	2.00	28.9
four-bond						
51	9	50	4	27	0.00	30.9
52	10	49	6	25	2.00	28.9
53	9	50	4	27	0.00	30.9
54, **55**[b]	10	49	6	25	2.00	28.9
five-bond						
56, **57**[b]	8	51	4	27	0.00	30.9
58	9	50	5	26	2.00	28.9
six-bond						
59	7	52	3	28	0.00	30.9

[a] For structures see Figure I, next page. [b] Pairs of isomorphic disconnections.

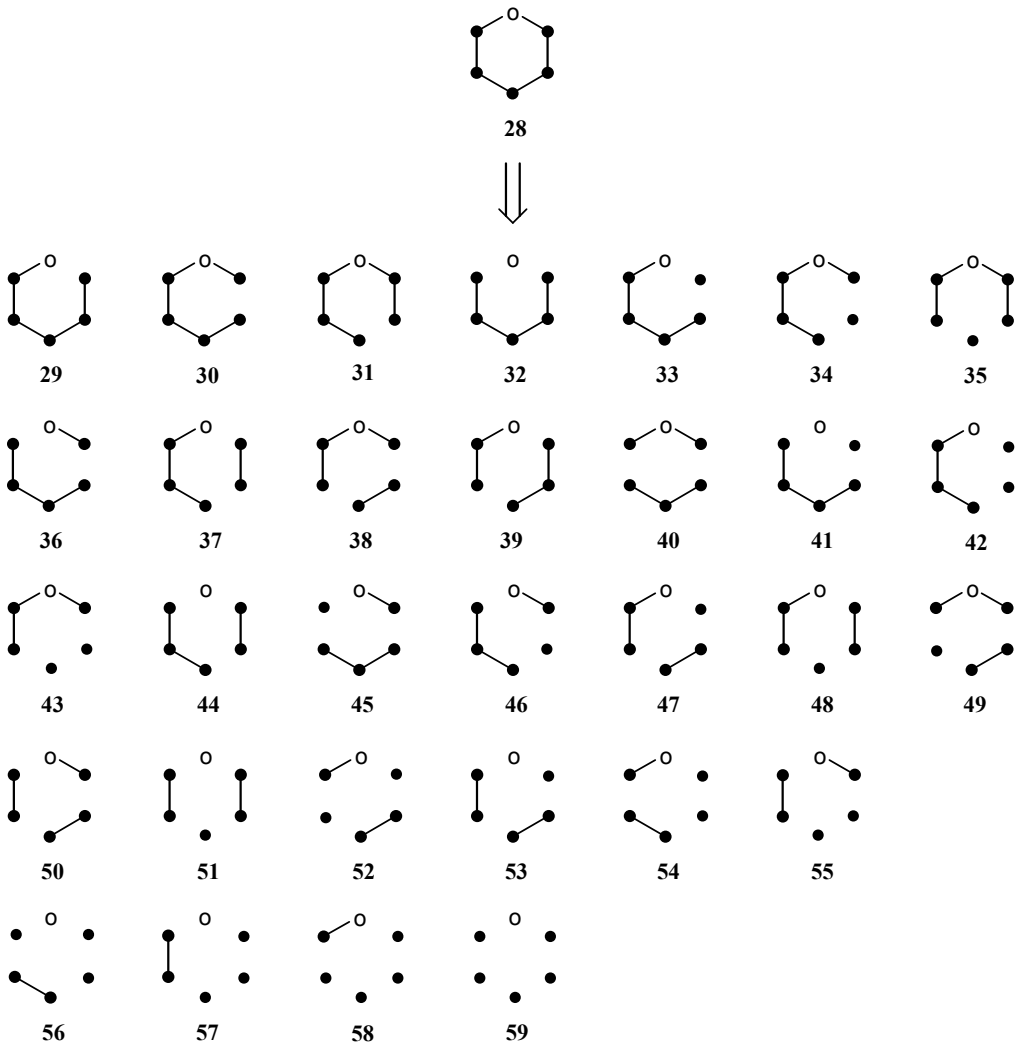

Figure I All possible disconnections of azacyclohexane (**28**).